

The following is a set of slides used on July 22, 1999 to conduct a workshop introducing the NSHAPC public use data files to potential federal agency and other users. They have been slightly modified to increase clarity in some instances. They should be read by anyone just beginning to use the NSHAPC files, as they give an overview/serve as an introduction that will save you a lot of time and confusion.

Throughout these slides, “UI” stands for “Urban Institute.” So “UI variables” means “variables created by Urban Institute analysts in the course of producing the NSHAPC reports, and expected to be of use to other researchers.”

Urban Institute
Washington, DC
August 1999

INTRODUCTION TO THE PUBLIC USE DATA FILES

**NATIONAL SURVEY OF HOMELESS
ASSISTANCE PROVIDERS AND CLIENTS
(NSHAPC)**

CONTENTS

- # NSHAPC design (pages 3 - 5)
- # NSHAPC data sources and how they relate to the public use tape file structure (pages 6 - 7)
- # File contents (pages 8 - 20)
- # Weights (pages 21-23)
- # UI variables and things to watch out for
- # Significance tests and the design effect

NSHAPC DESIGN

76 primary sampling areas

- the 28 largest MSAs
- 24 randomly sampled small and medium-sized MSAs
- 24 randomly sampled groups of rural counties or parts of counties

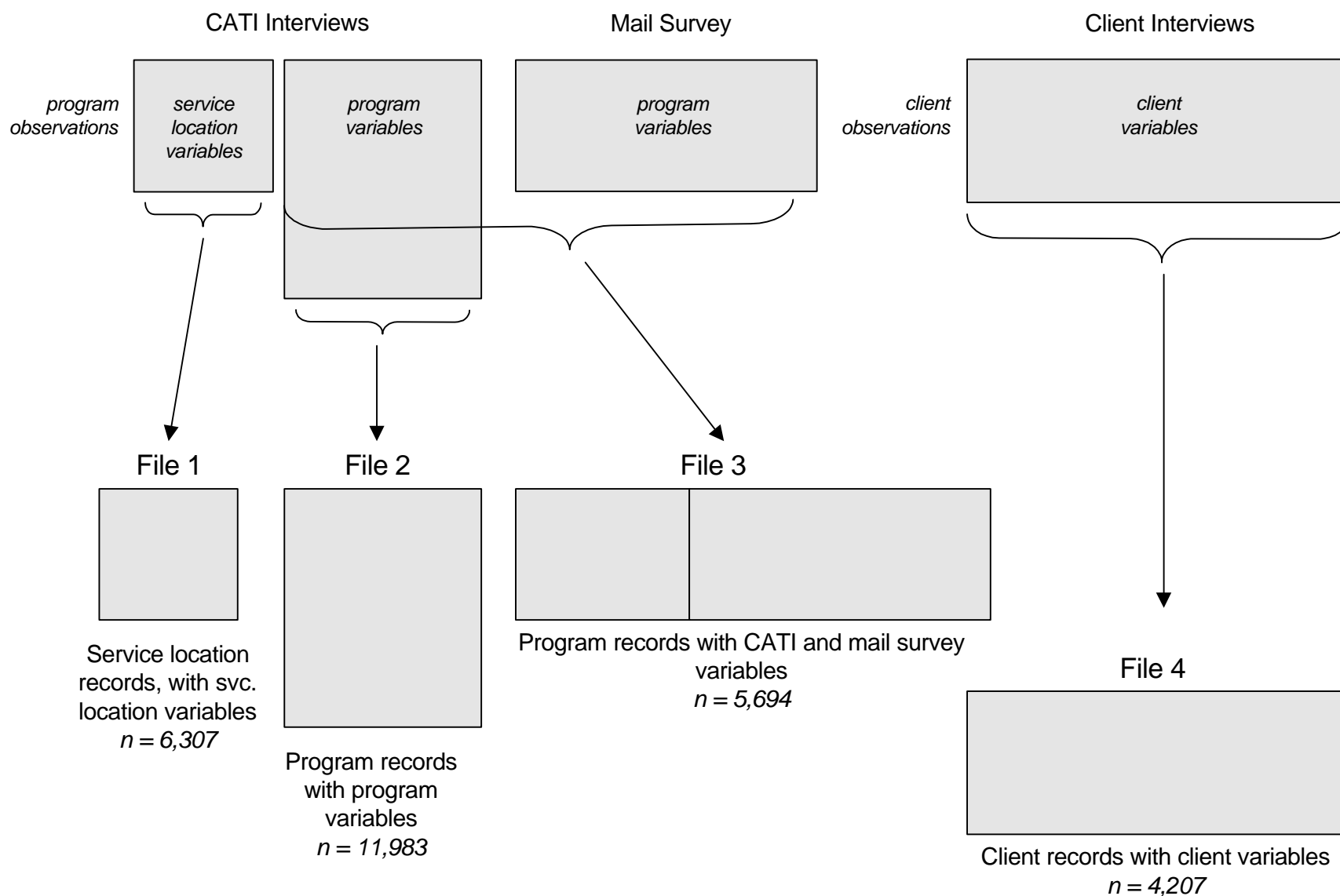
NSHAPC METHODS

- # Telephone survey (CATI) of service locations (providers)
 - Interviews with 6,307 service locations representing about 12,000 such locations nationwide
 - These locations reported operating 11,983 programs, representing about 40,000 such programs nationwide
- # Mail survey of programs, to get details about services
 - 5,694 valid programs ($UI_PRG \geq 1$)
- # Client interviews, in person with 4,207 clients
 - 4,133 from regular definition programs, 74 from revised definition programs

NSHAPC'S 16 PROGRAM TYPES

- # Emergency shelters
- # Transitional housing
- # Permanent housing for formerly homeless people
- # Voucher distribution for emergency accommodation
- # Acceptance of vouchers for emergency accommodation
- # Food pantries
- # Soup kitchens and meal distribution programs
- # Mobile food programs
- # Physical health care programs
- # Mental health care programs
- # Alcohol/drug programs
- # HIV/AIDS programs
- # Outreach programs
- # Drop-in centers
- # Migrant housing used in off-season for homeless shelter
- # Other

Relationship of Data Sources to File Structure



TIME PERIODS FOR DATA FILES

February 1996

- File 1: Service location data
- File 2: Program data from CATI for all programs for which $UI_PRG \geq 1$
- File 3: Program data from CATI and Mail Survey for all programs having a mail survey for which $UI_PRG \geq 1$

October-November 1996

- File 4: Client data

FILE 1: SERVICE LOCATION DATA

CATI data on File 1

- Geographic location (LOCALE, PSUTYPE)
- Programs offered at service location
 - Correspond to 16 original program types
 - Q2 i : “Do you offer an _____ program?”
- Primary mission of service location

UI variables on File 1

- Recode of primary mission to include ‘other’ responses in original or new categories
- UI_i i variables = recode of Q2 i (e.g., UI_ES, UI_SK)
- Number of UI_i i programs at a service location
- Recoded geographic location (URBRURAL)

DEFINING URBAN-RURAL STATUS

The urban-rural status variable, URBRURAL, was assigned using the variables LOCALE and PSUTYPE from the CATI

- Central city
 - PSUTYPE 1 or 2, and LOCALE = large or mid-sized central city
- Suburban/urban fringe
 - Program was in the urban fringe or a large or mid-size city; or the program was in large town, small town, or rural area and was a metropolitan statistical area (PSUTYPE 1 or 2 but not central city)
- Rural
 - PSUTYPE = 3. LOCALE could be large town, small town, or rural area, as long as it was outside an MSA

FILE 2: PROGRAM DATA FROM THE CATI- 1

CATI data on File 2

- Geographic location (LOCALE, PSUTYPE)
- Original program type assigned by Census (ORIG_PRG)
- Meals offered (breakfast, lunch, dinner), days open
- Clients the program expected to serve on an average day
 - By family type: 1- or 2-parent families, single adults, unaccompanied youth, children (topcoded at 301+)
 - Percent respondent estimated to be homeless
 - Percent female
- Type of sponsoring organization and funding sources
- Source of clients (where referrals come from)
- Population focus of program
 - Options included domestic violence, runaway/homeless youth, mental illness, alcohol/drug, mental illness and alcohol/drug, HIV/AIDS, veterans
- Receipt of USDA-donated foods; funded as part of Department of Veterans Affairs homeless programs

FILE 2: PROGRAM DATA FROM THE CATI- 2

UI variables on File 2

- Recoded population focus of program to include 'other' responses in original or new categories
- Recoded geographic location (URBRURAL)
- Region of the country
- Whether program is part of a UI-defined service location
- Program category, according to UI definition
 - standard and expanded categories (UI_PRG, UI_EXPRG)
- Total number of people the program expected to serve on an average day in February 1996
- Percent of people the program expected to serve on an average day whom the respondent thought were homeless

FILE 3: PROGRAM DATA FROM THE CATI AND THE MAIL SURVEY - 1

CATI data on File 3

- All variables in File 2

Mail survey data on File 3

- Focuses on client needs for 59 specific services in 11 general service groups:
 - food
 - clothing
 - life skills
 - case management
 - housing
 - education
 - employment
 - general health care
 - substance abuse
 - mental health
 - other: child care, DV counseling, legal assist., veterans' services

FILE 3: PROGRAM DATA FROM THE CATI AND THE MAIL SURVEY - 2

Mail survey data on File 3, (continued)

- For the 59 specific services, program respondent's perception of:
 - Proportion of clients who need the service
 - Frequency with which clients are able to get this need met
 - Whether, and where, the service is available in the area
- Transportation assistance and other services provided
- For housing programs only:
 - Maximum capacity for families, single persons
 - Average percent occupancy per season for families/single persons
 - Typical number of families/single persons turned away because of full occupancy per season
 - Reasons for turning families/single persons away
 - Reasons for operating at less than full capacity for families/single persons

- Destination of families/single persons leaving the program

FILE 3: PROGRAM DATA FROM THE CATI AND THE MAIL SURVEY - 3

- # UI variables on File 3 (described in more detail on pp. 15-20)
- All variables in File 2
 - In-scope interview status; flag for inclusion in UI analysis
 - Availability of specific service (transformed into 1 variable from 4)
 - Perceived need, per service group
 - Perceived availability of any service, per service group
 - Program category
 - Groups the 16 UI program types into the 4 categories of housing, food, health, other
 - Recoded need for specific service
 - Reverses values from 'All' equals 1 to 'All' equals 3, etc.
 - Recoded frequency with which need is met for specific service
 - Reverses values from 'Always' equals 1 to 'Always' equals 4
 - Precise availability of specific service
 - 'Some,' 'most,' or 'all' clients need this service
 - Values include 'only this program,' 'this program and other

offsite program,' etc.

IN-SCOPE STATUS FOR THE MAIL SURVEY

- # The in-scope status of a program - represented by the variable M_ISR_F - determined whether a program was used in UI analysis of the mail survey.
 - A program was in-scope if the observation met the UI definition of a program (non-missing UI_PRG) and was classified as an interview by Census (M_ISR = 1)

AVAILABILITY OF SERVICE - MAIL SURVEY

- # The AV_H### variables were created to summarize the availability of a specific service as reported on the mail survey.
- # Availability questions were only asked if respondent reported that 'all,' 'most,' or 'some' program clients need the service
 - Coded as **available** if respondent says that 'this program,' 'other on-site program,' or 'other off-site program' offer the service
 - Coded as **unavailable** in the area if the respondent says that the service is 'not provided anywhere.'
 - Information on the availability of the service is coded as **missing** if the respondent said that "no clients" need the service *or* did not answer the question 'who provides this service?'

SUMMARIZING RESPONSES ACROSS SERVICE GROUP - 1

- # The summary variables condense the mail survey information about the 59 specific services in 11 different service groups
- # Level of need: the variables S#0
 - Summarize the need level reported within the 11 service groups
 - 1 = At least some clients are reported to need at least one of the specific services in the service group (all items answered)
 - 2 = No client is reported to need any of the specific services in the service group (all items answered)
 - 3 = Some combination of 'no clients need this service,' 'don't know' and missing responses to all of the specific services in a service group (some items DK or Not Answered)
 - 99 = None of the 'proportion of clients who need this service' questions were answered for any of the specific services within a service group (all items Not Answered)

SUMMARIZING RESPONSES ACROSS SERVICE GROUP - 2

- # Service available at this program: the variables S#2
 - Summarize the provision of services at the respondent program in the 11 service groups
 - 1 = This program provides at least one of the specific services in the service group
 - 0 = None of the specific services in the service group are available at this program
 - 99 = Respondent did not answer the question 'who provides this service?' for any of the specific services in a service group *or* none of the program's clients need any of the specific services in a service group.

- # Similar logic was used to summarize the responses on
 - Service availability at another on-site program (the variables S#3)
 - Service availability at another off-site program (the variables S#4)

SUMMARIZING RESPONSES ACROSS SERVICE GROUP - 3

- # Service not provided anywhere: the variables S#5
 - Summarize the provision of services in the service group for whatever the respondent considered the program's geographical area
- 1 = At least one of the specific services in the service group is provided by this program, an on-site program, or an off-site program
- 2 = None of the specific services of the service group are provided anywhere
- 3 = Answers on some of the specific services in the service group were missing and others were not available in the area
- 99 = All responses on availability of specific services within the service group are missing

FILE 4: CLIENT SURVEY TOPICS

- # Current Living Condition
- # Living Situation for Those Clients Currently Without Regular Housing
- # Living Situation for Those Clients Currently With Regular Housing
- # Demographic Characteristics
- # Children and Education
- # Food Intake
- # Employment
- # Sources of Income and Service Use
- # Veteran Status
- # Current Physical Health
- # Victimization and Imprisonment
- # Mental Health
- # Chemical Dependency
- # Service Needs
- # Interviewer Observations

OVERVIEW OF NSHAPC WEIGHTS

Data Set	Weight	Weighted Findings Yield
File 1 CATI Service Location Data	CATIWGT	Nationally representative estimates of NSHAPC <i>service locations</i>
File 2 CATI Program Data	CATIWGT	Nationally representative estimates of NSHAPC <i>programs</i>
File 3 CATI & Mail Survey Data for Programs with a Mail Survey	MAILWGT	Nationally representative estimates of NSHAPC <i>programs</i>
File 4 Client Survey Data	CLIWGT	Nationally representative estimates of NSHAPC program <i>clients</i>

CLIWGT is the weight that should be used for all analyses for which the analyst wants to represent the universe of homeless assistance program clients in the country. The alternative, conducting analyses without using this weight, will produce only a description of the actual people who were interviewed and will not be representative of the country as a whole.

CLIWGT is calculated to reflect service use over the seven-day period preceding interviews of NSHAPC clients. It has been constructed to assure that a person is not double-counted even if she or he uses several homeless assistance programs during the course of that seven-day period. The seven-day period was selected as the most appropriate for several reasons. First, it takes advantage of the program use data collected from clients about a seven-day period. In doing so, it assures that no single day of idiosyncratic program use exercises an undue influence on the weight. Second, it recognizes and compensates for the fact that some people who are homeless on a given day might not be represented if they did not use a program on that day, but are more likely to be represented if a longer time frame is used. One does not want to underrepresent homeless clients who use services infrequently. Finally, it parallels the practice of using a seven-day weight first reported with respect to the 1987 Urban Institute study (Burt and Cohen, 1989).

ADJUSTMENTS TO WEIGHTS FOR TESTS OF STATISTICAL SIGNIFICANCE

- # Weights for analyzing data for service locations (CATIWGT) or programs (CATIWGT or MAILWGT) must be normalized to yield appropriate tests of statistical significance
- # “Normalizing” means reducing each weight so the sum of the weights equals the sample size
- # Normalize CATIWGT and MAILWGT by dividing each individual weight by the mean of that weight from the full sample
- # Client weight (CLIWGT) is already normalized, so you don’t have to do anything to it

APPROPRIATE & INAPPROPRIATE USES OF “FRAME” IN FILE 4: CLIENT DATA

- # The frame from which clients were sampled is a key element of the sampling design and construction of the client weights
- # However, its uses in analysis are limited
- # WHAT **NOT** TO USE “FRAME” FOR
 - DO NOT use FRAME to select a subsample of clients for analysis. If you do, your weights will be wrong.
 - If you want to look at the characteristics of people who use a particular type of program (e.g., soup kitchens), USE THEIR SURVEY RESPONSES TO DEFINE THE SUBSAMPLE (e.g., SOUPK)
- # WHAT IT IS OKAY TO USE “FRAME” FOR
 - If, using the whole client sample, you want to know where clients were found for the NSHAPC interview, look at “FRAME” (see p. 30 for another use)

WHAT IS A “PROGRAM”? — File 1

In File 1: Service Location Data

- $Q2i = 1$, Respondent says “yes,” we have a program of type i
- $UI_i = 1$ (e.g., UI_ES , UI_PH , UI_SK), Urban Institute considers that this service location has a program of type i

!!! Use UI_i , not $Q2i$!!!

- $Q2i$ and UI_i sometimes differ for several reasons:
 - ($Q2i > UI_i$) Respondent gave no further information about the program after the Q2 response
 - ($Q2i > UI_i$) For health programs, respondent said in answer to $Q2i$ R that the health program just named was part of a previously identified program
 - ($Q2i > UI_i$) Mail survey data indicated that the program was a duplicate, out-of-scope, only administrative or support in nature, really a different program type, closed, or never existed
 - ($UI_i > Q2i$) A write-in response to “Other” has been recoded
 - ($UI_i > Q2i$) Mail survey data indicated that the program was this type, but there was not a “yes” on its $Q2i$

WHAT IS A “PROGRAM”? — Files 2 and 3

In Files 2 and 3: Program Data

- UI_PRG = the program category into which UI researchers placed the program, based on 15 Census program types plus “Housing/Financial,” which was big enough in “Other” to make into its own category (range = 1-16, but some of the specific codes differ)
- UI_EXPRG= expanded program definitions created by UI, incorporating program speciality, population focus, primary mission (range = 1-106)
- ORIG_PRG = the program category into which Census originally placed the program, of Census’ 16 program types (range = 1-16) (DO NOT USE for analysis, only for understanding what changed from the CATI to the final program designations)

WHAT IS A “PROGRAM”? — File 4

In File 4: Client Data

- FRAME = the sampling frame (program type) from which the Census Bureau sampled the client (range = 1-12)

Program types OMITTED from the client sampling (because they were not likely to increase coverage of the homeless population significantly) were:

- Programs accepting vouchers
- All four types of health programs
- Outreach programs going TO homeless assistance programs (those going to the streets WERE sampled)
- Many programs classified by Census as “Other”

EXPANDED PROGRAM DEFINITIONS

(UIEX_PRG)

- # UI used information about a program's population focus if it only had one (UI_FOCi), **Primary** population focus (UI_PRIM) if it had more than one, and the service location's primary mission (MISSION) to identify a special focus for Emergency Shelters, Transitional and Permanent Housing Programs, and Soup Kitchens
- # If U_FOCi , UI_PRIM, and/or MISSION indicated a special program focus from among the following categories, UIEX_PRG reflects it:
 - MH
 - CD
 - MH/CD
 - HIV/AIDS
 - Domestic Violence (not done for PH or SK, insufficient N)
 - Youth “ ”
 - Family “ ”
- # You won't be able to reproduce this coding because MISSION is only on File 1 and UI_FOCi and UI_PRIM are only on Files 2 and 3.

REGULAR DEFINITION PROGRAMS AND REVISED DEFINITION PROGRAMS

Regular Definition Programs

- Meet the survey criteria for a homeless assistance program
- The most important of these is “having a FOCUS on serving homeless people”

Revised Definition Programs

- Are found only in rural areas
- Too few programs in rural areas met the regular definition
- Relaxed the criterion of “having a FOCUS on serving” to just plain “SERVING” homeless people

Variable is REVDEF (0 = regular program; 1 = revised definition program)

WHEN THE NSHAPC REPORTS USE REGULAR AND REVISED DEFINITION PROGRAMS AND/OR THEIR CLIENTS

- # All analyses in the Summary and Technical NSHAPC reports are based on programs with a non-missing UI_PRG. Most of these are regular definition programs (REVDEF = 0), but some are revised definition programs (REVDEF = 1)
- # All analyses in these NSHAPC reports covering SERVICE LOCATION or PROGRAM data INCLUDE revised definition programs
- # All analyses in the NSHAPC reports covering CLIENT data EXCLUDE clients found in revised definition programs
- # You may do whatever you want with REVDEF, but be sure to read the appendix to Chapter 13 of the Technical Report so you make an informed decision about what to include or exclude when analyzing client data

TIME FRAME PROBLEMS WHEN ANALYZING PROGRAM USE

- # Information sources for housing program use
 - Q1.1a -- Where live today? (Day of interview = 8th day)
 - Q2.1a,b and Q3.1a,b -- What used during past 7 days
 - Sampling frame where found (8th day)
 - Q2.2a,b and Q3.2a,b -- Ever used, most recent use

- # Information sources for food program use
 - Q6.6a,b -- What used during past 7 days
 - Sampling frame where found (8th day)
 - Q6.6c,d -- Ever used, most recent use

- # Information sources for outreach and drop-in center use
 - Q2.3a,b and Q3.3a,b -- What used during past 7 days
 - Sampling frame where found (8th day)
 - Q2.3c,d, Q3.3c,d -- Ever used, most recent use

- # ISSUE -- THESE SOURCES DO NOT ALWAYS AGREE

DEFINING HOMELESSNESS STATUS

- # A client was classified as ***currently homeless*** (HOMLSS=1) if he/she:
- Reported staying in any of the following places on the day of the survey or during the 7-day period prior to being interviewed: an emergency shelter, a transitional housing program, a hotel or motel paid for by a shelter voucher, an abandoned building, a place of business, a car or other vehicle, or anywhere outside.
 - Reported that the last time they had a place of their own for 30 days or more in the same place was more than 7 days ago.
 - Said their last period of homelessness ended within the last 7 days.
 - Was identified for inclusion in the NSHAPC client survey at an emergency shelter or a transitional housing program, or at a voucher distribution program, but only if there was at least one other indicator of current homelessness.
 - Reported getting food from “the shelter where you live” within the last 7 days.
 - Reported staying in their own or someone else’s place on the day of the interview but said they “could not sleep there for the next month without being asked to leave.”

- # A client was classified as ***formerly homeless*** (HOMLSS=2) if he/she did not meet any of the conditions to qualify as currently homeless, but reported that any of the following occurred during their lifetime:
- Reported staying in any of the following places: an emergency shelter, a transitional housing program, a welfare/voucher hotel, an abandoned building, a place of business, a car or other vehicle, anywhere outside, or a permanent housing program for the formerly homeless.
 - Reported a previous period of homelessness.
- # The remainder of NSHAPC clients were classified as never homeless (HOMLSS=3). They are referred to as “other service users” throughout the summary and technical reports.

DEFINING SERVICE USE PATTERN

SOURCES OF INFORMATION FOR FAMILY STATUS

Demographics Section

- Is anyone with you or are you by yourself?
- Are you with your... Spouse
Partner/boyfriend/girlfriend
Children
Other relatives
Other persons?

Children and Education Section

- *For each child [you have] under age 18,*
Does [child's name] live with you?

DEFINING FAMILY STATUS

Client is classified as in a “Family” if:

- # In residential settings (shelters and housing programs)
 - S/he reported being with child(ren) now, AND
 - S/he reported living with at least one of her/his own children under age 18

- # In all other settings
 - S/he reported living with at least one of her/his own children under age 18

Anyone not classified as a “Family” is classified as “Single”

DEFINING MENTAL HEALTH PROBLEMS

Past Month MH

- Scoring at or above 0.25 on ASI, or
- Treated for emotional/mental problems within past month, or
- Took prescribed medications for psychological/emotional problems within past month, or
- Identified “mental health condition” as most important thing keeping them homeless, or
- Any history of treatment AND at least one ASI “condition” within past month

Past Year MH

- Past Month MH, or
- Same criteria as for past month, but within past year

Lifetime MH

- Past Year MH, or
- Same criteria as for past year, but for lifetime, or
- Stayed in a psychiatric hospital or a group home for mentally ill (during homeless spell)

DEFINING ALCOHOL PROBLEMS

Past Month A

- Scoring at or above 0.17 on ASI, or
- Treated for alcohol abuse within past month, or
- Drank to get drunk three or more times a week within past month, or
- Any history of treatment AND drank three or more times a week within past month

Past Year A

- Past Month A, or
- Same criteria as for past month, but within past year

Lifetime A

- Past Year A, or
- Same criteria as for past year, but for lifetime, or
- Three or more alcohol-related difficulties in lifetime (from MAST)

DEFINING DRUG PROBLEMS

Past Month D

- Scoring at or above 0.10 on ASI, or
- Treated for drug abuse within past month, or
- Use drugs intravenously (from current physical health section), or
- Used any of a variety specific illegal drugs three or more times a week within past month

Past Year D

- Past Month D, or
- Same criteria as for past month, but within past year

Lifetime D

- Past Year D, or
- Same criteria as for past year, but for lifetime, or
- Three or more drug-related difficulties in lifetime (from DAST)

DEFINING ADM PROBLEMS

Past Month ADM

- Past Month A, D, or M, or
- Reported “addiction to alcohol or drugs” as most important thing keeping them homeless

Past Year ADM

- Past Year A, D, or M

Lifetime ADM

- Lifetime A, D, or M, or
- Stayed in a residential recovery program (during homeless spell)

STATISTICAL TESTS AND THE DESIGN EFFECT

- # A Census Bureau average design effect for NSHAPC is **3**
- No matter what type of statistical test you are doing, you must use this design effect as part of your formula to get the correct level of statistical significance
 - See handout for the formulas we used for confidence intervals and comparison of percentages. In regression analyses, multiply the standard error for each coefficient by 3 before conducting significance tests.
 - If you are dealing with a subset of the data and that subset is likely to be particularly skewed with regard to where it was found (e.g., Native Americans or Hispanics), the design effect is probably even greater (possibly as much as double), so either use a higher *alpha* level (e.g., .01), interpret with caution, or

both.

HOW TO GET PUBLIC USE FILES

FREE

- GO TO www.census.gov
- SELECT Subjects A to Z
- SELECT “N” and then look for “NSHAPC”
- Download data files, codebook files, special instruction files
- Everything you need to know is in those files, please read them first, second, and third

FILES ON CD, for a charge

- Call Census Bureau Customer Service at 301-457-4100